

SADA: Improved Data Symbolization and Optimization Method on HAR from Microscopic Perspective

*

1st Huichao Men
School of Computer Science and Engineering
Northeastern University
Shenyang, P.R.China
candyji@sina.com

2nd Botao Wang
School of Computer Science and Engineering
Northeastern University
Shenyang, P.R.China
wangbotao@cse.neu.edu.cn

Abstract—Nowadays, human activity recognition(HAR) becomes a hot topic with broad applications. Some researches have conducted HAR from microscopic perspective and achieved good results. In this article, two methods are proposed for further improvement. Firstly, an improved symbolization method with stacked sparse autoencoder is proposed for better data symbolization. Secondly, an improved multi-classification Adaboost is proposed to further optimize the recognition effect, and it is more suitable for the application scenario of this article. In the experiments section, firstly, experiments and analysis about various influencing parameters are conducted, then comparison experiments with several new or representative methods are carried out, and finally five representative sensor activity datasets(UCI Sports and Daily dataset, Wisdm Phoneacc&Watchacc dataset, Skoda dataset, HAPT dataset) are used to prove the universal applicability and achieve satisfactory effect.

Index Terms—Human Activity Recognition, Data Symbolization, Ensemble Learning, Machine Learning

I. INTRODUCTION

Human activity recognition is one of the hottest research topics nowadays. In this article, we mainly conduct research on sensor data. In some previous researches [1] [2], human activity recognition combining with NLP models from microscopic perspective has been proved meaningful in the recognition of variable length activities, and it has achieved good effect. Figure 1 is an example of decomposing one activity from microscopic perspective. In order to further improve the effect, we conduct deeper research on data symbolization about these sub-segments, and Adaboost is improved for better recognition effect, too.

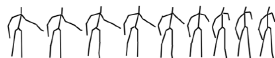


Fig. 1. One Macroscopic Activity Decomposed to a Microscopic Sequence

In the previous research, the main problem is that the data symbols in an activity sequence are messy and the



(a) Messy Symbols Line (b) Regular Symbols Line

Fig. 2. Comparison Between Messy and Regular Symbols

regularity is weak. Data symbol plays a core role in the process of migrating from NLP to HAR. The messy symbols are equivalent to noise points, which is negative on activity sequence recognition. The logical and regular symbols have an positive effect on recognition. Therefore, some researches made some efforts on feature extraction and different symbols to improve regularity. However, it still has limited effect on improving sequence recognition.

Figure 2 is a significant comparison between messy and regular activity symbols in a activity sequence.

Figure 2a shows that a messy symbol connection line like noise polyline, almost without regularity. However, the regular symbol connection line (Figure 2b) is more smooth, it is easier to recognize. The horizontal axis represents the serial numbers of points in a sequence and the vertical axis represents is the symbol of one certain point. For example, point 25 to 35 in Figure 1a can be symbolized as [e,e,b,e,b,f,e,d,d,a,a], while the same part in Figure 2b is [d,d,b,b,b,b,b,b,b]. It is obvious that the symbol sequence in Figure 1b is more regular, concise, and easier to recognize.

The symbol points are mainly converted from the data sub-segments through feature extraction, so a good feature extraction method [3] beneficial to recognition is very important. Sparse autoencoder(SAE) is a kind of neural network that is usually used in feature extraction. Since the activity data collected by sensor is noisy, and the sparse autoencoder is more suitable for extracting features from the noisy data, a feature extraction method based on stacked SAE with good

effect is proposed in this article.

Since the obtained data symbols are regular and logical, like natural language, many NLP recognition models could be migrated to HAR. In order to improve the effect of basic recognition models, Adaboost [4], an ensemble learning model is used in this article. It is usually used to optimize other machine learning models. In this article, it is further improved to adapt to the application scenario studied in this article.

In this paper, these contributions are proposed:

- (1) Propose a new unsupervised human activity data symbolization method to improve the symbol regularity and logic, including stacked SAE with L-BFGS and clustering symbolization algorithm. In-depth theoretical analysis of the choice of clustering algorithm is conducted, too.
- (2) An improved Adaboost with multi-classification is proposed. Two weight-setting choices are discussed.
- (3) These two parts above form a complete model. Compared with the previous research, the new model has a better recognition effect.
- (4) Comparison experiments of different parameters and verification experiments on multiple datasets are conducted, proving the effectiveness and universal applicability of the proposed structure. LSTM is one of the most common models applied to sequence classification and recognition. Therefore, it is used as a tool to verify the two parts of method proposed in this article.

II. RELATED WORK

Data symbolization is one of the research focuses in this article. Feature extraction is the main part of it. In many articles, statistical features [5] [6], time and frequency features [7] [8], deep learning feature extraction method (such as convolution feature [9]), etc, are the popular feature extraction methods. They all do not perform best in the research of this article. Covariance matrices [10], sliding window, enlarging window, PCA, etc, are methods aiming to expand the amount of information. Sparse autoencoder is a feature extraction model based on neural network and it is suitable for noise data. In this article, it is improved to extract better features.

Adaboost, a sequence ensemble learning method, is usually used as a tool to optimize various machine learning models. In this article it is used to optimize activity recognition models. Many classic multi-classification Adaboost models [4] convert a binary-classification into a multi-classification, such as Adaboost.MH, Adaboost.M1, and Adaboost.M2. They all have disadvantages. SAMME is a multi-classification Adaboost with unsatisfactory stability [11]. GAdaboost is a state-of-the-art method with higher effectiveness but lower accuracy [12].

III. PRIOR KNOWLEDGE

In some previous researches [2], activities are decomposed into sub-segments and converted to data symbols with relevance and logic. This logical relationship is similar to that in natural language. Therefore, models from NLP could be migrated to human activity recognition.

Activity-unit is the smallest unit in the activity recognition. Original activity data is decomposed to sub-segments, then converted to activity-unit with feature extraction and symbolization. It is similar to "alphabet". **Activity-combination** represents a group of related combined activity-units with semantic meaning and it reflects the relevance and logical relationship between sub-segments. It is similar to "word". **Activity-sequence** is a sentence composed of activity-combinations, which is similar to "sentence".

Activities are converted to these forms of data similar to natural language. The logical relationship of natural languages is migrated to human activity recognition field. In our research, we will conduct further exploration based on it.

IV. METHODOLOGY

Figure 3a is the overview of the whole model in this article, named **SADA**. It means that SAE and Adaboost are the most important components in this model. The upper part is the overall process of it. The two big boxes are detailed figures of the data symbolization method and the improved Adaboost method. Details will be introduced in this section.

Firstly, the original training/test data is converted to data symbols (activity-units) through the improved data symbolization method. These activity-units compose regular and logical activity-sentences. Then, training activity-sentences are trained with the model optimized by improved Adaboost. Finally, the test activity-sentences are recognized by the trained model.

A. Human Activity Data Symbolization Method Based on Unsupervised Multi-layer Stacked SAE

As shown in Figure 3b, the structure in this section consists of two parts: stacked SAE (combining with L-BFGS) and clustering symbolization method. Because the sub-segments in activities are all unlabeled data, unsupervised feature extraction algorithm is necessary to generate data symbols (activity-units).

Feature extraction is the core part of data symbolization in this section. Because it is necessary to use unsupervised algorithm to generate data symbols, the output layer (such as softmax layer) in common application of SAE is removed. Data symbols are converted from feature data with clustering/classification algorithms.

In practical applications, the feature extraction effect of single-layer SAE is unsatisfactory. This article proposes a deep learning method of multiple layers stacked SAE to extract more proper features of activity data sub-segments layer by layer. As shown in Figure 3b, multiple SAEs are cascaded in multiple layers. The feature data $h_{i,j}$ of the previous layer is transferred to the next layer $h_{i+1,j}$ to continue to extract features. Finally the most representative features are obtained and the dimension is reduced. L-BFGS is used as the parameter optimization algorithm in this method. Through this method, the original data is converted to more representative **feature data**. The theoretical analysis of numbers of layers is described in Section 4.

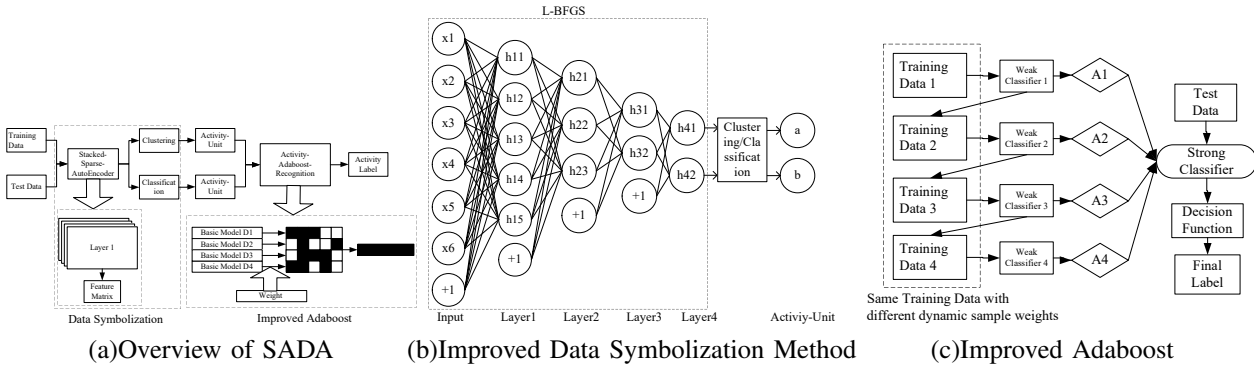


Fig. 3. Structure of SADA

Then, the feature data is converted to activity-unit(data symbol) with clustering algorithm. The principle is to generate diversified activity-units and minimize the amount of scattered points(noisy points). Classification symbolization algorithm usually corresponds to the clustering algorithm which has been chosen. Theoretical analysis is described in Section 4.3.

The activity-units generated in this part compose **activity-sequences**. They are very concise, regular and logical, which are suitable for human activity recognition.

B. An Improved Adaboost for Optimizing the Human Recognition Model

After the symbolization operation, the regular and concise activity-units converted from the activity data will be applied to the recognition models migrated from NLP. In this section, Adaboost is used to optimize these basic recognition models. Because of its decision function, basic Adaboost is usually used for binary-classification. It is not suitable for multi-classification. Therefore, improving the decision function is feasible to convert the basic Adaboost to a multi-classification Adaboost.

Figure 3c shows the process of the improved Adaboost in this section.

Here is the decision function of basic Adaboost. It is a binary classifier.

$$Final_Classifier(x) = \text{sign}\left(\sum_{n=1}^N A_n Weak_Classifier_n(x)\right)$$

A is the weak classifier weights generated from iterative training.

In this section, multiple same activity recognition models with dynamic sample weights are set as the weak classifiers. Then, they are integrated into a stronger classifier. In experimental exploration, it is found that although the average recognition accuracies of some weak classifiers are declined, the recognition of different categories have achieved complementary effects, so the overall recognition accuracy can be improved.

The method in this section includes two parts. Primary Adaboost weight and fine-tuning weight could both obtain good accuracy, and the fine-tuning one is better adapt to the

actual application scenarios of this article. Primary weight is a weight array generated from iterative training without fine-tuning.

In the training period, the weak classifiers are trained in the form of **iteration and sequence concatenation**. Weak classifiers and primary weight array, A , are obtained in this period.

1) *Improved Adaboost with Primary Weight*: In decision function of test period, a group of weak classifiers' results based on test data have been obtained. Then it is converted to an one-hot matrix(Formula(3)).

This group of formulas is the basic decision function in this article:

$$A = [a_0, a_1, \dots, a_{n-1}] \quad (2)$$

$$Sub_Prediciton_Onehot = [[0, \dots, 1], \dots, [0, \dots, 1]] \quad (3)$$

$$Result_Matrix = A \times Sub_Prediction_Onehot \quad (4)$$

$$Result_Matrix_Sum = \text{sum}(Result_Matrix[\text{row}]) \quad (5)$$

$$Final_Label = \text{max_index}(Result_Matrix_Sum) \quad (6)$$

A is the Adaboost primary weight array of weak classifiers. $Sub_Prediciton_Onehot$ is the one-hot matrix of weak classifiers' recognition results. It is a $n * m$ matrix. n is the number of weak classifiers, setting as 4 here. m is the number of activity categories. $Result_Matrix$ is the cross product of A and $Sub_Prediciton_Onehot$. $Result_Matrix_Sum$ is an array which is the result of $Result_Matrix$ **added row by row**. $Final_Label$ is the maximum index of it.

2) *Improved Adaboost with Fine-tuning Weight*: Adaboost with primary weight has achieved good results. By analyzing the experiment data, in order to further improve the recognition effect, a fitting function is applied in this part.

Through the analysis of the parameters generated by multiple weak classifiers based on three datasets (UCI Dataset, Wisdm Phoneacc/Watchacc Dataset), an **average empirical constant scale array** of the primary weight can be estimated. Since this constant scale array is just estimated and only shows the trend, it does not represent the most appropriate weight ratio for each dataset. Therefore, we adjust the trend of estimated ratio and primary weight to fine-tune the weak classifiers weight array.



Fig. 4. An Example of the Trend of Transformed Primary Weight and the Tangent Function Based on Three Datasets

The image of the relationship between the **transformed** primary weight array and average constant scale array looks like the trend of tangent function(Figure 4). The final fine-tuning weight is adjusted by approximate function to make it close to the estimated ratio and suitable for the actual situation of each dataset. We try to use the transformation of the tangent function as the form of fine-tuning function.

According to the characteristic of the primary weight in basic Adaboost, it more depends on the distribution form of the data. Since the activity types collected by most common sensor activity datasets are similar, the data distribution patterns are also similar. Therefore, the trend of average scale array in many activity datasets are relatively stable. Therefore, the fitting method here is not just an empirical method, it can be applied in many datasets and has obtained acceptable results. Tangent function is just one choice of fitting function. Maybe there are other functions suitable, too.

In fact, the average constant scale is just a tool for analysis and it is not applied in actual calculation in this model.

In Figure 4, three polylines represent the weight trend of three datasets, **which is just shown as a similarity expression of the trend**. The original weight line has been **transformed** into the form in this figure by **stretching and rotating** without changing the trend of the turning point. The only one curve is the standard trend of tangent function. The vertical line represents the average scale array. The horizontal line represents the weight axis after stretched and rotated operation. The center cross lines are to assist in marking the coordinate axis of the tangent function. It does not indicate the real position of the coordinate axis of three polylines and is only a formal representation. Therefore, this figure shows that the trends of the primary weight of the three datasets all conform to the tangent trend.

$$AN = \text{normalize_sort}(A) \quad (7)$$

$$k = \text{degrees}(\arctan(\frac{AN_0 - AN_{n-1}}{2})) \quad (8)$$

$$AW_i = \tan(\frac{\pi * k}{max} * AN_i) + \frac{AN_0 + AN_{n-1}}{2} \quad (9)$$

Formula(7)(8)(9) compose a fitting function group to fine-tune primary weight. *max* and *min* are the maximum and minimum value of the primary weight array. *AW* is the fine-tuning weight array. *n* is set as 4 here, too. Degree function converts radians to angles. Based on the fine-tuning weight, the *A* in Formula 1 can be replaced by *AW*.

Here are the total steps of the improved Adaboost in this section:

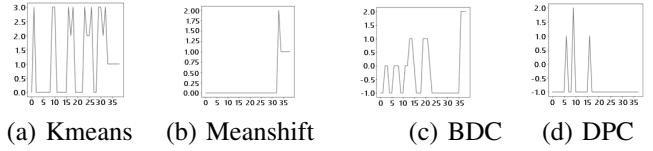


Fig. 5. Data Symbolization with Different Clustering Algorithms

- (1) The whole recognition model is trained with the concise and regular training activity-sequences based on dynamic sample weights. Primary weight array *A* and weak classifiers are obtained after iterative training.
- (2) If primary weight is chosen for the improved Adaboost in this section, go to Step(4), otherwise go to Step(3).
- (3) *A* is converted to *AW* through the fitting function group.
- (4) One test activity-sequence converted by the improved data symbolization is tested in weak classifiers and the recognition results group is converted to an one-hot matrix. The *Result_Matrix* is the cross product of weight array and one-hot matrix.
- (5) The final label is the index of the maximum of the array which is the result of *Result_Matrix* added row by row.

After this process, the final label of the activity is obtained. The time complexity of the improved Adaboost is in the same order of magnitude as the basic Adaboost.

C. Analysis about Selection of Clustering and Classification Symbolization Method

Clustering algorithm is an important part of data symbolization. In this article, the selection principle is to ensure the diversity of activity-units while minimizing the number of scattered points. Figure 5 shows the difference between different algorithms.

Center-based and density-based clustering algorithms are the most commonly clustering algorithms. Meanshift and KMeans are two representative center-based algorithms. As shown as Figure 5b, because Meanshift does not set the clustering number, it clusters most points into one or few clusters, which makes it difficult to separate different activities and to distinguish between activities because of rare kinds of points(activity-units), although it looks concise. And KMeans most meets the selection principle as shown as Figure 5a. High parameter-tuning requirement makes it easy for density-based algorithms to produce scattered points(labeled as -1), as Figure 5c. DPC, a state-of-the-art density-based algorithm, faces the same problem in Figure 5d.

Finally, KNN is chosen for classification because of its corresponding algorithm, KMeans.

D. Analysis about Sub-segment Length for Data Symbolization & Number of Stacked SAE Layers

Sub-segment Length for Data Symbolization: It is an important parameter in data symbolization. For NLP sequence recognition, a longer sequence contains more logical information between activity-units and it is beneficial to recognition.

TABLE I
COMPARISON EXPERIMENTS OF DIFFERENT SUB-SEGMENT LENGTHS
AND DIFFERENT NUMBERS OF STACKED SAE LAYER

Length	10	5	3
Average Acc	0.707	0.802	0.858
QC Numbers	7	12	14
TOP1 Acc	0.987	1	1
Layer	1	3	4
Average Acc	0.766	0.853	0.858
QC Numbers	12	13	14
TOP1 Acc	1	1	1

In order to obtain a longer sequence, the length of sub-segment must be shorter for a constant length of data.

Number of Stacked SAE Layers: The original SAE is just a single-layer structure, which cannot meet the requirement of this research. The multi-layer structure is more conducive to feature extraction, but after reaching a certain value, the increasing layer number have little effect on feature extraction, and the computational cost will increase.

V. EXPERIMENTS

Experiments in this section are taken on 5 sensor datasets, including UCI Sports and Daily Dataset(19 categories) [13], WISDM Phoneacc&Watchacc Dataset(18 categories) [6], HAPT Dataset(6 categories) [14] and Skoda Dataset(10 categories) [15]. UCI dataset is applied in Section 5.1&5.2 &5.3. Skoda dataset is applied in Section 5.3.

In this section, WISDM Dataset is a composition structure of "combination of main categories and sub-categories". Distinction between main categories and the similarity between sub-categories is very obvious. It is more suitable for two-layer recognition model, so a set of single body sample is just for demonstration here.

"QC Numbers" in Table 1&2 represents "the number of categories with accuracy over 0.8".

In this section, LSTM is chosen as the migrated NLP recognition model to verify the proposed method. According to the previous research, LSTM is a proper tool to recognize concise and regular sequence with balanced efficiency and accuracy. Other NLP model may be proper, too, but it is not the focus in this article.

A. Comparison Experiments of Different Sub-segment Length & Different Numbers of Stacked SAE Layer

This part of the experiments in Table 1 verifies the parameter analysis in Section 4.4. From the upper part of Table 1, it can be known that a small and proper length value is more suitable for the HAR in this research. From the lower part of Table 1, it shows that enough and proper layers are beneficial to feature extraction and activity recognition. 4-layers-structure achieve the best, but 3-layers-structure is also satisfactory. A 4-layer structure is used in later experiments.

TABLE II
ABLATION EXPERIMENTS

Methods	WL	PL	CL	DL
Average Acc	0.708	0.382	0.784	0.802
QC Numbers	8	3	9	13
TOP1 Acc	1	1	1	1
Methods	DL	DML	SADA	
Average Acc	0.802	0.809	0.858	
QC Numbers	13	13	13	
TOP1 Acc	1	1	1	

B. Progressive Ablation Experiments and Comparison Experiments

The progressive ablation experiments include two parts. Firstly, since the core part of data symbolization is feature extraction. Two latest feature extraction methods(WISDM Feature [6] and PSD [8]) and classical convolve feature [9] are compared with the proposed data symbolization in this article(in Table 2). The recognition tool is simple LSTM.

Secondly, Adaboost.M2 [11], a classical multi-classification model, is compared with simple LSTM and the improved Adaboost proposed in this article.

In Table 2,(1)WL=WISDM Feature+LSTM (2)PL=PSD Feature+LSTM (3)CL=Convolve Feature+LSTM (4)DL=Proposed Data Symbolization in this article+LSTM (5)DML=Proposed Data Symbolization in this article with Adaboost.M2+LSTM (6)SADA=Proposed Whole Model in this article. According to the conclusions in article [1], these methods are suitable for comparison.

In this section, the ablation experiments prove the effectiveness and advantages of the two parts of the proposed structure. WISDM Feature and PSD Feature both do not archive satisfactory effects. The method based on convolution has been greatly improved, but it is still poorer than the proposed data symbolization method. Simple recognition model(LSTM) without optimization and Adaboost.M2 are also less effective than the proposed optimization method in this article.

C. Additional Comparison Experiments

The experiments in last section have verified that the two parts of the method proposed in this article has a better effect based on previous research.

ILVote [16] is a latest human activity recognition model based on incremental learning and vote. Transformer [17] is one of the hottest models nowadays in various research fields. to prove the proposed method in this article. BERT [18] is a latest model developed from Transformer, mainly used for NLP. In this section, they are taken as comparisons **without feature extraction** to prove the advantage of our whole proposed method.

The main shortcoming of ILVote is highly dependent on manual parameter setting, especially the full activity window length. In many cases, it is difficult to determine a proper full activity window length. Therefore, although the theoretical basis of ILVote is meaningful, it is lack of universal applicability.

TABLE III
ADDITIONAL COMPARISON EXPERIMENTS

UCI	DL	SADA	ILVote	Transformer	BERT
Average Acc	0.802	0.858	0.250	0.279	0.175
QC Numbers	13	13	2	2	2
TOP1 Acc	1	1	1	0.927	0.98
Skoda	DL	SADA	ILVote	Transformer	BERT
Average Acc	0.730	0.797	0.233	0.108	0.0547
QC Numbers	5	8	1	0	0
TOP1 Acc	1	1	0.85	0.407	0.395

TABLE IV
VERIFICATION EXPERIMENTS BASED ON DIFFERENT DATASETS

Primary Weight	UCI	WPA	WWA	SK	HA
AA	0.857	0.528	0.364	0.745	0.738
Fine-tuning Weight	UCI	WPA	WWA	SK	HA
Increase Categories	3	4	6	5	2
Decrease Categories	1	0	3	1	1
AA	0.858	0.553	0.488	0.797	0.754

Because of difficulty to set a proper window length, the results are unsatisfactory.

In the absence of proper feature extraction in this part of experiments, the training and test sequences imported into the Transformer and BERT are very irregular and lack of logic. Even this state-of-the-art models can not recognize them well. Therefore, it verifies that the concise and regular sequence generated by the proposed method in this article is meaningful.

As shown as the experiments results in Table 3, after obtaining regular and logical sequence, LSTM, a most common NLP model, could achieve good recognition effects.

D. Verification Experiments with Primary Weight and Fine-tuning Weight

This section is a group of verification experiments based on the whole structure proposed in this article with primary weights and fine-tuning weights on 5 datasets. In Table 5(next page):(1)UCI=UCI Sports and Daily Dataset (2)WPA=WISDM Phoneacc Dataset (3)WWA=WISDM Watchacc Dataset (4)SK=Skoda Dataset (5)HA=HAPT Dataset (6)FT=Fine-Tuning (7)AA=Average Accuracy

From the Table 4, the proposed model has achieved good results with primary weights. In case of model with fine-tuning function on some datasets, although the accuracy of a small numbers of categories decrease, more categories' accuracy increase, and the average accuracy increase, too. For example, in the experiments results of Skoda and HAPT, only one activity's accuracy decreases, and the average accuracy and the accuracies of most activities all increase.

Experiments in this section prove the universal applicability and effectiveness of the proposed structure both with primary weights and fine-tuning weights.

VI. CONCLUSION AND FUTURE WORK

Analyzing activity from microscopic perspective with integrated NLP methods is the foundation of our research. In this

article, we propose a new data symbolization method and an improved Adaboost for basic recognition models with good effect and universal applicability. Both of them need be further improved. In the future, we will continue to take more research on this topic and try to further improve the recognition effect.

VII. ACKNOWLEDGEMENT

This work is supported by "the Fundamental Research Funds for the Central Universities"(N2104002).

REFERENCES

- [1] Huichao Men, Botao Wang, Gang Wu, "Mitar: A study on human activity recognition based on NLP with microscopic perspective (in press)," *Frontiers of Computer Science*, 2020.
- [2] Mengyan Yuan, Enqing Chen, and Lei Gao, "Posture selection based on two-layer ap with application to human action recognition using hmm," in *IEEE International Symposium on Multimedia*, 2017.
- [3] Isabelle Guyon and Andr Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, no. 6, pp. 1157–1182, 2003.
- [4] Ji Zhu, Ann Arbor, and Trevor Hastie, "Multi-class adaboost," *Stats & Its Interface*, vol. 2, no. 3, pp. 349–360, 2006.
- [5] Dorra Trabelsi, Samer Mohammed, Yacine Amirat, and Latifa Oukhelou, "Activity recognition using body mounted sensors: An unsupervised learning based approach," in *International Joint Conference on Neural Networks*, 2018.
- [6] Gary M. Weiss, Kenichi Yoneda, and Thair Hayajneh, "Smartphone and smartwatch-based biometrics using activities of daily living," *IEEE Access*, 2019.
- [7] Marcin Grzegorzec, Michael Reinhold, and Heinrich Niemann, "Feature extraction with wavelet transformation for statistical object recognition," in *Computer Recognition Systems, Proceedings of the 4th International Conference on Computer Recognition Systems, CORES'05, May 22-25, 2005, Rydzyna Castle, Poland*, 2005.
- [8] Xiao Li, Yufeng Wang, Bo Zhang, and Jianhua Ma, "Psdrrn: An efficient and effective har scheme based on feature extraction and deep learning," *IEEE Transactions on Industrial Informatics*, vol. PP, no. 99, pp. 1–1, 2020.
- [9] Pradeep Pujari Mohit Sewak, Md.Rezaul Karim, *Practical Convolutional Neural Networks: Implement advanced deep learning models using Python*, China Machine Press, 2019.
- [10] Hamza Ergezer and Kemal Leblebicio?Lu, "Time series classification with feature covariance matrices," *Knowledge & Information Systems*, 2017.
- [11] YUAN Shun YANG Xinwu, MA Zhuang, "Multi-class adaboost algorithm based on the adjusted weak classifier," *Journal of Electronics & Information Technology*, 2016.
- [12] Mai F. Tolba and Mohamed Moustafa, "Gadaboost: Accelerating adaboost feature selection with genetic algorithms," in *ECTA, 8th International Conference on Evolutionary Computation Theory and Applications*, 2016.
- [13] Kerem Altun, Billur Barshan, and Orkun Tunel, "Comparative study on classifying human activities with miniature inertial and magnetic sensors," *Pattern Recognition*, vol. 43, no. 10, pp. 3605–3620, 2010.
- [14] Jorge L Reyes-Ortiz, Luca Oneto, Albert Sama, Xavier Parra, and Davide Anguita, "Transition-aware human activity recognition using smartphones," *Neurocomputing*, vol. 171, pp. 754–767, 2016.
- [15] P Zappi, T Stiefmeier, E Farella, and D Roggen, "Activity recognition from on-body sensors by classifier fusion: sensor scalability and robustness," in *International Conference on Intelligent Sensors*, 2007.
- [16] Hengnian Qi, Kai Fang, Xiaoping Wu, Lili Xu, and Qing Lang, "Human activity recognition method based on molecular attributes," *International Journal of Distributed Sensor Networks*, vol. 15, no. 4, pp. 155014771984272, 2019.
- [17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, "Attention is all you need," *arXiv*, 2017.
- [18] Jinhuyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang, "Biobert: a pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, Sep 2019.